Planar unclustered scale-free graphs as models for technological and biological networks.

Alicia Miralles, Francesc Comellas*

Departament de Matemàtica Aplicada IV, Universitat Politècnica de Catalunya
08860 Castelldefels, Catalonia, Spain

Lichao Chen, Zhongzhi Zhang

School of Computer Science and Shanghai Key Lab of Intelligent Information Processing,
Fudan University, Shanghai 200433, China

Abstract

Many real life networks present an average path length logarithmic with the number of nodes and a degree distribution which follows a power law. Often these networks have also a modular and self-similar structure and, in some cases - usually associated with topological restrictions- their clustering is low and they are almost planar. In this paper we introduce a family of graphs which share all these properties and are defined by two parameters. As their construction is deterministic, we obtain exact analytic expressions for relevant properties of the graphs including the degree distribution, degree correlation, diameter, and average distance, as a function of the two defining parameters. Thus, the graphs are useful to model some complex networks, in particular several families of technological and biological networks, and in the design of new practical communication algorithms in relation to their dynamical processes. They can also help understanding the underlying mechanisms that have produced their particular structure.

Key words: complex networks, scale-free networks, self-similar graphs, modular graphs, planar graphs.

*Corresponding author. PMT Edifici C3, c/ Esteve Terradas 5, 08860 Castelldefels, Barcelona, Catalonia, Spain,Tel. +34 934 134 109 Fax +34 934 137 007

Email addresses: `[almirall,comellas]@ma4.upc.edu` (Alicia Miralles, Francesc Comellas), `zhangzz@fudan.edu.cn` (Lichao Chen, Zhongzhi Zhang)

1. Introduction

Ten years have past since the publication of the groundbreaking papers by Watts and Strogatz [1] on small-world networks and Barabasi and Albert [2] on scale-free networks. Their works led researchers to the design of new network models to describe complex systems in nature and society like the Internet, protein-protein interactions, transportation systems or social and economic networks. Their models try to match observational studies which have identified at least three important common characteristics for real-life networks: They exhibit a small average distance and diameter (compared to a random network with the same number of nodes and links); the number of links attached to the nodes obeys a power-law distribution (the networks are scale-free); and recently it has been discovered that, often, real networks are self-similar [3] and show a degree hierarchy related to the modularity of the system, see [4, 5, 6] .

Many of the proposed models are stochastic as this is the case for the now classical preferential attachment method [2]. Thus, the use of mean field techniques is required to estimate the main parameters of a network [7]. However, a deterministic approach has proven useful to complement and enhance the probabilistic and simulation techniques. Deterministic models have a clear advantage, as they allow an analytical exact determination of relevant network parameters, which then can be compared with experimental data coming from real and simulated networks

Among the different methods known to generate deterministic models those based on recursive or iterative methods are of particular interest. In these methods, new nodes are added and connected to a given substructure of the network at each generation step. This is the case for pseudo-fractal networks [8] where, at each step, new vertices are added simultaneously, one to each already existing link. This construction can be generalized if complete subgraphs of a given size (cliques) are considered instead of links (which are of course 2-cliques), see [9]. Similar rules give the Apollonian networks [10, 11, 12]. On the other hand, there also exist techniques that produce networks by duplication of a given substructure, see [13, 14].

A generalization of these two methods introduces at each iteration a substructure which is added to the network, according to a deterministic rule. Substructures that have been used are triangles [15], cycles [16] and paths [17].

In this paper we go one step further by considering the simultaneous

introduction of d substructures in parallel -in our case, paths- which are attached to the same basic unit (a link) generalizing the model given in [17], which added a single path to each link. The resulting graphs are essentially different from those in [17]. In particular they are scale-free (with a power-law exponent which depends on d) while in [17] the degree distribution is exponential. The model is a family of planar, modular, hierarchical and self-similar networks, with small-world scale-free characteristics and with clustering coefficient zero, and all these parameter are determined by d as well as by the iteration step t. We note that some important real life networks, for example those associated to electronic circuits, Internet and some biological systems [18, 19], have these characteristics as they are modular, almost planar and with a reduced clustering coefficient and have small-world and scale-free properties. Thus, these networks are modeled by our construction which can be considered as a new tool in the study of their associated complex systems. In particular, the model could be used to find also practical algorithms in relation to dynamical processes (synchronization, cover time, etc.) for these technological and biological networks and can help understanding the underlying mechanisms that have produced their particular structure.

In the next section we introduce the family of graphs object of study and in Section 3 we calculate analytically some relevant properties for the graphs, namely, the degree distribution, degree correlations, the diameter and the average distance. The last section provides some conclusions.

2. Generation of the graphs $M_d(t)$

In this section we introduce a family of modular, self-similar and planar graphs which have the small-world property and are scale-free. The family depends on an adjustable parameter d and the iteration number t. We provide an iterative algorithm, and also a recursive method, for its construction. The construction methods allow a direct determination of the number of vertices (nodes) and edges (links) of the graph.

Iterative construction.– We give here an iterative formal definition of the proposed family of graphs, $M_d(t)$, characterized by $t \geq 0$, the number of iterations and a parameter d associated with the self-repeating modular structure.

First, we call *generating edges* the only edge of $M_d(0)$ and all edges of $M_d(t)$ whose endvertices have been introduced at different iteration steps t.

All other edges of $M_d(t)$ will be known as *passive edges.* A generating edge becomes passive after its use in the construction.

The graph $M_d(t)$ is constructed as follows:

For $t = 0$, $M_d(0)$ has two vertices and a generating edge connecting them.

For $t \geq 1$, $M_d(t)$ is obtained from $M_d(t-1)$ by adding, to every generating edge in $M_d(t-1)$, d parallel paths of length three (each path has four vertices and three edges) by identifying the two final vertices of each path with the endvertices of the generating edge.

The process is repeated until the desired number of vertices is reached, see Fig. 1. We note that the number of vertices can be also adjusted with the parameter d (number of parallel paths that are attached to each generating edge).

Recursive modular construction.– The graph $M_d(t)$ can also be defined as follows:
(a) For $t = 0$, $M_d(0)$ has two vertices and a generating edge connecting them.
(b) For $t = 1$, $M_d(1)$ is obtained from $M_d(0)$ by adding to its only edge d parallel paths of length three by identifying the two final vertices of each path with the endvertices of the initial edge.
(c) For $t \geq 2$, $M_d(t)$ is made from $2d$ copies of $M_d(t-1)$, by identifying, vertex to vertex, the initial edge of each $M_d(t-1)$ with the generating edges of $M_d(1)$, see Fig. 1.

Number of vertices and edges of $M_d(t)$.– We use the following notation: $\tilde{V}(t)$, $\tilde{E}(t)$ and $\tilde{E}_g(t)$ denote, respectively, the set of vertices, edges and generating edges introduced at step t, while $V(t)$ and $E(t)$ denote the set of vertices and edges of the graph $M_d(t)$.

Notice that, at each iteration, a generating edge is replaced by $2d$ new generating edges and d passive edges. Therefore: $|\tilde{E}_g(t+1)| = 2d \cdot |\tilde{E}_g(t)|$, and $|\tilde{E}_g(t)| = (2d)^t$. As each generating edge introduces at the next iteration $2d$ new vertices and $3d$ new edges we have $|\tilde{V}(t+1)| = 2d \cdot |\tilde{E}_g(t)| = (2d)^{t+1}$ and $|\tilde{E}(t+1)| = 3d \cdot |\tilde{E}_g(t)| = 3d \cdot (2d)^t$. As $|\tilde{V}(0)| = 2$ and $|\tilde{E}_g(0)| = 1$, the number of vertices and edges of $M(t)$, $t \geq 0$, is:

$$\begin{aligned} |V(t)| &= \sum_{i=0}^{t} |\tilde{V}(i)| = \frac{(2d)^{t+1} + 2d - 2}{2d - 1}, \\ |E(t)| &= \sum_{i=0}^{t} |\tilde{E}(i)| = \frac{3d(2d)^t - d - 1}{2d - 1}. \end{aligned} \tag{1}$$

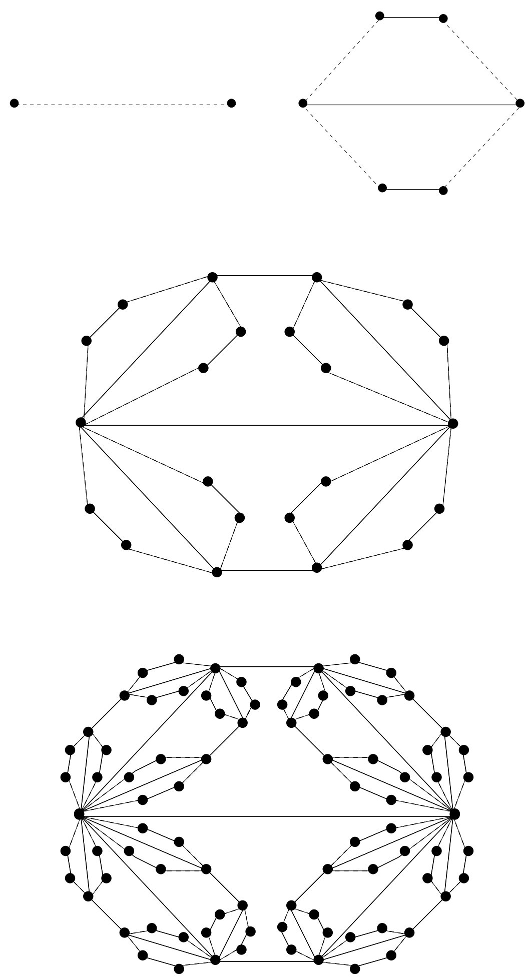

Figure 1: Graphs $M_d(t)$ produced at iterations $t = 0, 1, 2$ and 3 for $d = 2$.

Planarity.– A graph is planar if it can be drawn on the plane with no edges crossing. By construction of $M_d(t)$, the introduction at each iteration of d parallel paths connected to each generating edge, which afterwards becomes passive, adds $2d$ new vertices to the graph and they can be drawn without crossing edges. Planarity could also be proven from Kuratowski's theorem or from the known planarity test which states that a graph is planar if it has no cycles of length 3 and $|E| \leq 2|V| - 4, |V| > 3$, see [20].

3. Topological properties of $M_d(t)$

Thanks to the deterministic nature of the graphs $M_d(t)$, we can give exact values for the relevant topological properties of this graph family, namely, the degree distribution, degree correlations, the diameter and the average distance.

Degree distribution.– Initially, at $t = 0$, the graph has two vertices of degree one. When a new vertex i is added to the graph at iteration t_i, this vertex has degree 2 and it is connected to only one generating edge. We use the following notation: $k_g(i,t)$, $k_p(i,t)$ and $k(i,t)$ are, respectively, the number of generating edges, passive edges and total edges connected to vertex i, at step $t \geq t_i$. Therefore $k(i,t) = k_g(i,t) + k_p(i,t)$ is the degree of vertex i at this step.

From the construction process we can write,

$$\begin{cases} k_p(i,t+1) = k_p(i,t) + k_g(i,t) \\ k_g(i,t+1) = dk_g(i,t) \end{cases} \tag{2}$$

with the initial conditions,

$$\begin{aligned} k_g(i,t_i) &= 1, \quad t_i \geq 0 \quad \text{and} \\ k_p(i,t_i) &= \begin{cases} 0 & \text{if} \quad t_i = 0 \\ 1 & \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

and for $d > 1$ we have,

$$\begin{aligned} k_g(i,t) &= d^{t-t_i}, \quad t_i \geq 0 \quad \text{and} \\ k_p(i,t) &= \begin{cases} 1 + \frac{d^t - d}{d-1} & \text{if} \quad t_i = 0 \\ 2 + \frac{d^{t-t_i} - d}{d-1} & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

All the vertices that have been introduced at step t_i have the same degree at step t:

1. The two vertices introduced at step $t_i = 0$ have degree,

$$\begin{aligned} k(i,t) &= k_g(i,t) + k_p(i,t) = \\ &= d^t + 1 + \frac{d^t - d}{d-1} = \frac{d^{t+1}-1}{d-1}. \end{aligned} \tag{5}$$

2. The $|\tilde{V}(t_i)| = (2d)^{t_i}$ vertices introduced at step $t_i > 0$ have degree,

$$\begin{aligned} k(i,t) &= k_g(i,t) + k_p(i,t) = \\ &= d^{t-t_i} + 2 + \frac{d^{t-t_i} - d}{d-1} = 1 + \frac{dd^{t-t_i}-1}{d-1}. \end{aligned} \tag{6}$$

Therefore the degree spectrum of the graph is discrete and to relate the exponent of this discrete degree distribution to the power law exponent of a continuous degree distribution for random scale free networks, we use the technique described by Newman in [19] to find the cumulative degree distribution $P_{\text{cum}}(k)$. If we denote by $V(t,k)$ the set of vertices that have degree k at step t,

$$\begin{aligned} P_{\text{cum}}(k) &= \frac{\sum_{k' \geq k} |V(t,k')|}{|V(t)|} = \frac{2 + \sum_{t_i'=1}^{t_i} (2d)^{t_i'}}{\frac{(2d)^{t+1}+2d-2}{2d-1}} = \\ &= \frac{(2d)^{t_i+1} + 2d - 2}{(2d)^{t+1} + 2d - 2} = \\ &= \frac{(2d)^{t - \frac{\ln(k + \frac{2-k}{d} - 1)}{\ln(d)} + 1} + 2d - 2}{(2d)^{t+1} + 2d - 2}. \end{aligned}$$

Fot t large, we obtain,

$$\begin{aligned} P_{\text{cum}}(k) &\approx (2d)^{-\frac{\ln(k + \frac{2-k}{d} - 1)}{\ln(d)}} = (k + \frac{2-k}{d} - 1)^{-\frac{\ln(2d)}{\ln(d)}} \\ &= k^{-\frac{\ln(2d)}{\ln(d)}} (1 - \frac{1}{d} + \frac{2-d}{kd})^{-\frac{\ln(2d)}{\ln(d)}}. \end{aligned}$$

For $k >> 1$ this expression gives

$$P_{\text{cum}}(k) \approx k^{-\frac{\ln(2d)}{\ln(d)}} (1 - \frac{1}{d})^{-\frac{\ln(2d)}{\ln(d)}} \tag{7}$$

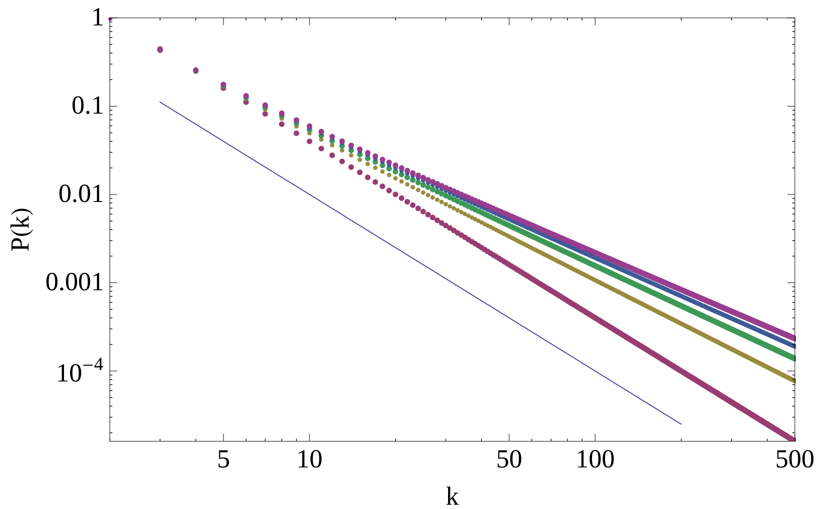

Figure 2: Log-log representation of the cumulative degree distribution for $M_d(t)$, $d = 2, 3, 4, 5$. The reference line has slope -2.

Thus, the degree distribution follows a power-law

$$P_{\rm cum}(k) \sim k^{-\gamma}, \quad \text{with } \gamma = \frac{\ln(2d)}{\ln(d)},$$

and therefore the degree distribution is scale-free, see Fig. 2.

Research on networks associated to electronic circuits show that many of them are almost planar, modular and have a small clustering coefficient and in most cases their degree distributions follow a power-law [18, 19] with exponent values in the same range than those of $M_d(t)$.

Correlation coefficient.– We have obtained the Pearson correlation coefficient [21], $r(d,t)$, for the degrees of the endvertices of the edges of $M_d(t)$. In Appendix A we present the details of the calculation that leads to the its exact analytical expression as shown in Eq. 11. We particularize this general analytical result for different instances of the graphs, obtaining numerical values of the correlation as shown in Table 1 .

From the analytical results and the numerical values of the correlation coefficient we see that this family of graphs has the degrees of the endvertices negatively correlated (large degree vertices tend to be connected with low

	$t=1$	$t=2$	$t=3$	$t=10$
$d=2$	-0.1667	-0.0886	-0.0460	-0.0003
$d=10$	-0.4091	-0.2338	-0.1174	-0.0009
$d=100$	-0.4901	-0.2057	-0.0934	-0.0007

Table 1: Correlation coefficient at steps $t = 1, 2, 3, 10$ for several values of d.

degree vertices) and the graphs are disassortative, as it occurs with many technological and biological networks [19].

For $d >> 1$, we obtain $r(d,t) \approx \frac{1}{1 - 3 \cdot 2^{t-1}}$, which for t large gives $r(d,t) \sim 0$.

Diameter.– At each iteration step we introduce, for every generating edge, $2d$ new vertices. These vertices are among them at distance at most 3. As each vertex joins the graph of the former step through one new edge, the diameter will increase by exactly 2 units. Therefore $D(t) = D(t-1) + 2$, $t \geq 2$. As $D(1) = 3$, we have that the diameter of $M_d(t)$ is $D(t) = 3 + 2 \cdot (t-1)$, $t \geq 1$. Therefore, from Eq. 1, and as for t large, $t \sim \ln |V(t)|$ we have in this limit that $D(t) \sim \ln |V(t)|$.

Average distance.– The average distance of $M_d(t)$ is defined as:

$$\bar{D}(t) = \frac{1}{|V(t)|(|V(t)|-1)/2} \sum_{i,j \in V(t)} d_{i,j} \, , \tag{8}$$

where $d_{i,j}$ is the distance between vertices i and j.

In Appendix B we use the modular recursive construction of $M_d(t)$ to calculate the exact value of $\bar{D}(t)$ which results:

$$\begin{aligned} \bar{D}(t) &= (-1 + 4d - 5d^2 + 2d^3 + 2^{1+t} d^{1+t} - 7 \cdot 2^{2t} d^{2+2t} + \\ &+ 3 \cdot 2^{1+2t} d^{3+2t} - 2^{1+t} d^{1+t} t + 3 \cdot 2^{1+t} d^{2+t} t - \\ &- 2^{2+t} d^{3+t} t - 2^{1+2t} d^{2+2t} t + 2^{2+2t} d^{3+2t} t) \\ &/ ((-1 + 2d)(-1 + d + 2^t d^{1+t})(-1 + 2^{1+t} d^{1+t})). \end{aligned} \tag{9}$$

Notice that for a large iteration step, $t \to \infty$, $\bar{D}(t) \simeq t \sim \ln |V(t)|$, which shows a logarithmic scaling of the average distance with the number

of vertices of the graph. As we have a similar behavior for the diameter, the graph is small-world.

4. $M_d(t)$ as a model for some technical and biological networks

The graphs introduced here have parameters which are similar to those of some real life networks. A good example is the largest benchmark considered in [18] –a network with 24097 nodes, 53248 edges, average degree 4.34 and average distance 11.05– has a degree distribution which follows a power-law with exponent 3.0, and it has a small clustering coefficient $C = 0.01$ and other network properties are also in the same range than those of the graph $M_6(4)$, see [19]. Table 2 compares some network parameters from instances of our model with data coming from real networks published elsewhere. Although there are many similarities between the two sets, the aim of this model is not to match perfectly all the network parameters for some real life complex systems, but to provide an analytical framework where to perform precise tests of new algorithms (routing, synchronization, etc.) and check properties that otherwise would require less general and precise techniques like simulation of stochastic methods.

Network	Vertices	Edges	γ	Avg. dist.	Clust.	Avg. degree	r	Ref(s).
$M_2(5)$	1366	2047	3	6.850	0	2.997	-0.001	
Java Dev. Fram.	1376	2174	2.5	6.39	0.06	3.160	-0.002	[22, 19]
$M_6(2)$	158	235	2.39	3.290	0	2.975	-0.233	
Silwood Pk food web	154	366	1	3.4	0.15	4.75	-0.31	[23]
$M_6(3)$	1886	2827	2.39	4.474	0	2.998	-0.130	
protein inter. S.C.	2115	2240	2.4	6.80	0.071	2.089	-0.156	[24, 19]
$M_6(4)$	22622	33931	2.39	5.557	0	3.000	-0.007	
electronic circuits	24097	53248	2.39	11.05	0.01	4.34	-0.130	[18, 19]
$M_8(3)$	4370	6553	2.33	4.482	0	2.999	-0.123	
power grid	4941	6594		19.99	0.1	2.669	-0.003	[1, 19]

Table 2: Some instances of $M_d(t)$ and possible real network counterparts.

5. Conclusion

The graphs $M_d(t)$ introduced and studied here are planar, modular, have a disassortative degree hierarchy and are small-world and scale-free. Another

relevant characteristic of the graphs is their zero clustering. A combination of a low clustering coefficient, modularity, and small-world scale-free properties can be found in some real networks, in particular in technological and biological networks [19, 18], and most of them are also disassortative.

Finally, we should emphasize that the planar property and the deterministic character of the family, in contrast with more usual probabilistic approaches, should facilitate the exact determination of other network parameters and the development of new network algorithms that then might be extended to real-life complex systems.

Acknowledgments

F. Comellas and A. Miralles are supported by the Ministerio de Ciencia e Innovación, Spain, and the European Regional Development Fund under project MTM2008-06620-C03-01 and partially supported by the Catalan Research Council under grant 2009SGR1387. L. Chen and Z. Zhang are supported by the National Natural Science Foundation of China under Grant No. 60704044.

APPENDICES

A. Correlation coefficient calculation.

The Pearson correlation coefficient, $r(d,t)$, for the degrees of the endvertices of the edges of $M_d(t)$ is:

$$r(d,t) = \frac{|E(t)|\sum_i j_i k_i - [\sum_i \frac{1}{2}(j_i + k_i)]^2}{|E(t)|\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [\sum_i \frac{1}{2}(j_i + k_i)]^2} \tag{10}$$

where j_i, k_i are the degrees of the endvertices of the ith edge, with $i = 1, \cdots, |E(t)|$, see [21] .

To calculate the correlation coefficient we need to know the degree distribution of the endvertices of the edges in $\tilde{E}(t_i)$ at a given step t_i. We denote by $\langle j,k \rangle$ an edge connecting vertices of degrees j and k.

The detail of this distribution is given as follows:

The edges introduced at step t_i are:

1. Edges $\langle 2,2 \rangle$, connecting two vertices introduced at step $t_i > 0$. There are $(2d)^{t_i}/2$ edges (a half of the vertices introduced at step t_i). Notice that there is one edge $\langle 1,1 \rangle$ introduced at $t_i = 0$.

2. Edges $\langle 2, k(i', t_i)\rangle$ connecting vertices of degree two, introduced at step t_i, with all the vertices i' introduced at step $t_{i'}$ with $0 \leq t_{i'} \leq t_i - 1$. For each vertex i' there are $k_g(i', t_i)$ edges:
From the two vertices introduced at $t_{i'} = 0$, see (5), there are $2d^{t_{i'}}$ edges $\langle 2, \frac{d^{t_i+1}-1}{d-1}\rangle$.
From the $(2d)^{t_{i'}}$ vertices introduced at $t_{i'} > 0$, see (6), there are $(2d)^{t_{i'}} d^{t_i - t_{i'}}$ edges $\langle 2, 1 + \frac{d\,d^{t_i - t_{i'}}-1}{d-1}\rangle$.

Table 3 here displays a summary of the results.

Step t_i	Edges at step t_i	Number	Edges at step $t > t_i$
$t_i = 0$	$\langle 1, 1\rangle$	1	$\langle \frac{d^{t+1}-1}{d-1}, \frac{d^{t+1}-1}{d-1}\rangle$
$1 \leq t_i \leq t$	$\langle 2, 2\rangle$	$\frac{(2d)^{t_i}}{2}$	$\langle 1 + \frac{dd^{t-t_i}-1}{d-1}, 1 + \frac{dd^{t-t_i}-1}{d-1}\rangle$
	$\langle 2, \frac{d^{t_i+1}-1}{d-1}\rangle$	$2d^{t_i}$	$\langle 1 + \frac{dd^{t-t_i}-1}{d-1}, \frac{d^{t+1}-1}{d-1}\rangle$
$2 \leq t_i \leq t$ $1 \leq t_{i'} \leq t_i - 1$	$\langle 2, 1 + \frac{dd^{t_i - t_{i'}}-1}{d-1}\rangle$	$(2d)^{t_{i'}} \cdot d^{t_i - t_{i'}}$	$\langle 1 + \frac{dd^{t-t_i}-1}{d-1}, 1 + \frac{dd^{t-t_{i'}}-1}{d-1}\rangle$

Table 3: Number of edges in $M_d(t)$ according to the degrees of their endvertices.

Using these results, we can find the following sums:

$$
\begin{aligned}
\sum_i j_i k_i \;=\;& (4 + 16\,d - 51\,d^2 + 41\,d^3 - 8\,d^4 - \\
-\;& 3\,d^5 + d^6 + d^{t+1}(40 + 8t - 80 \cdot 2^t) + \\
+\;& d^{t+2}(-184 - 40t + 282 \cdot 2^t) + \\
+\;& d^{t+3}(306 + 74t - 373 \cdot 2^t) + \\
+\;& d^{t+4}(-236 - 64t + 227 \cdot 2^t) + \\
+\;& d^{t+5}(86 + 26t - 63 \cdot 2^t) +
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad d^{t+6}(-12-4t+7\cdot 2^t)+\\
&+ \quad d^{2t+2}(10+4t)+d^{2t+3}(-43-18t)+\\
&+ \quad d^{2t+4}(62+28t)+d^{2t+5}(-35-18t)+\\
&+ \quad d^{2t+6}(6+4t))/((d-1)^3\\
&\quad\;\; (2d^2-5d+2)(d-2)),
\end{aligned}
$$

$$
\begin{aligned}
\sum_i (j_i+k_i) \;&=\; \frac{-2}{(2d-1)(d-2)(d-1)^2}(-2-3d+\\
&+ \quad 10d^2-6d^3+d^4+d^{t+1}(-4+16\cdot 2^t)+\\
&+ \quad d^{t+2}(14-37\cdot 2^t)+d^{t+3}(14+26\cdot 2^t)+\\
&+ \quad d^{t+4}(4-5\cdot 2^t)-d^{2t+2}+3d^{2t+3}-\\
&- \quad 2d^{2t+4}),
\end{aligned}
$$

$$
\begin{aligned}
\sum_i (j_i^2+k_i^2) \;&=\; -2(-8-32d+186d^2-282d^3+\\
&+ \quad 145d^4+49d^5-96d^6+48d^7-11d^8+\\
&+ \quad d^9+d^{t+1}(-72+160\cdot 2^t)+d^{t+2}(384-\\
&- \quad 728\cdot 2^t)+d^{t+3}(-750+1252\cdot 2^t)+\\
&+ \quad d^{t+4}(606-882\cdot 2^t)+d^{t+5}(-33-39\cdot 2^t)+\\
&+ \quad d^{t+6}(-285+456\cdot 2^t)+d^{t+7}(201-286\cdot 2^t)+\\
&+ \quad d^{t+8}(-57+74\cdot 2^t)+d^{t+9}(6-7\cdot 2^t)+\\
&+ \quad 4d^{3t+2}-16d^{3t+4}+17d^{3t+5}+5d^{3t+6}-\\
&- \quad 19d^{3t+7}+11d^{3t+8}-2d^{3t+9})/((d^2-2d+1)\\
&\quad\;\; (2d-1)(d-2)(d^3-2d^2-2d+4)(d-1)^2).
\end{aligned}
$$

Replacing these sums into equation (10) we obtain, for any d, the exact analytical expression for the Pearson correlation coefficient of $M_d(t)$ which is displayed as Eq. 11. For $d=2$ this equation becomes Eq. 12:

B. Analytical determination of the average distance.

The average distance of $M_d(t)$ is defined as:

$$
\bar{D}(t) = \frac{1}{|V(t)|(|V(t)|-1)/2} \sum_{i,j\in V(t)} d_{i,j}\,, \tag{13}
$$

$$r(d,t) = \frac{(d(3\cdot 2^t d^t - 1) - 1)\cdot (B(d,t) + C(d,t)) - (D(d,t) + E(d,t))^2}{(d(3\cdot 2^t d^t - 1) - 1)\cdot F(d,t) - (D(d,t) + E(d,t))^2}, d > 2 \tag{11}$$

Where:

$$\begin{aligned}
B(d,t) &= (d^3 + 2\cdot d^2 - 6d - 1) + d^{t+1}2(2d-1)((5-3d) + \\
&+ t(1-d)) + d^{t+2}(-28\cdot 2^t + (-5+2t)d^t) + d^{t+3}(7\cdot 2^t + (6+4t)d^t), \\
C(d,t) &= (d^{t+3}(31\cdot 2^t + (-11+6t)d^t) + d^{t+2}(-90\cdot 2^t + \\
&+ (10-12t)d^t) + d^{t+1}80\cdot 2^t)/((d-2)^2), \\
D(d,t) &= d^2 - 3d - 1 + d^t(6\cdot 2^t - 6d^t) + d^{t+1}(11\cdot 2^t - 3d^t - 2) + d^{t+2}(-5\cdot 2^t - 2d^t + 4), \\
E(d,t) &= -12d^t(d^t - 2^t)/(d-2), \\
F(d,t) &= (d^{5+t}(-6 + 7\cdot 2^t + 2d^{2t}) + d^{4+t}(21 - 32\cdot 2^t + d^{2t}) + d^{3+t}(3 + 3\cdot 2^t - d^{2t}) + \\
&+ d^{2+t}(-42 + 62\cdot 2^t) + d^{1+t}(18 - 40\cdot 2^t) - d^5 + 5d^4 - 5d^3 - 11d^2 + 14d + 2)/(d^2-2).
\end{aligned}$$

For $d = 2$:

$$r(2,t) = \frac{4^t t^2 - 2^{t+1}t + 3\cdot 2^{2t+1}t - 2^{3t+2}t + 13\cdot 4^t - 3\cdot 2^{t+1} + 4^{2t+1} - 3\cdot 2^{3t+2} + 1}{2^{4t+1}t^2 + 2^{2t+1}t - 2^{3t+3}t + 2^{4t+3}t - 2^t + 5\cdot 4^t - 2^{3t+4} + 3\cdot 2^{4t+3} - 3\cdot 2^{5t+2}} \tag{12}$$

where $d_{i,j}$ is the distance between vertices i and j. In what follows, $S(t)$ will denote the sum $\sum_{i,j\in V(t)} d_{i,j}$.

The modular recursive construction of $M_d(t)$ allows us to calculate the exact value of $\bar{D}(t)$. At step t, $M_d(t+1)$ is obtained from the juxtaposition of $2d$ copies of $M_d(t)$, which we label $M_{d,t}^{(\eta)}$, $\eta = 1, 2, \cdots, 2d$, see Figures 1 and 3. Whenever possible, we drop the subscript d and represent $M_{d,t}^{(\eta)}$ as $M_t^{(\eta)}$ to keep the notation uncluttered. The copies are connected one to another at $2d+2$ vertices which we call *connecting vertices*. Two of them are the initial two vertices of the graph, which will be denoted in this section as A and B.

In Fig. 3 we display A, B and four more of these vertices, denoted as u, v, w and x. Note that in this figure, and for the sake of clarity, each copy of $M_d(t)$ has been represented as a rectangle, and only its connecting vertices have been drawn.

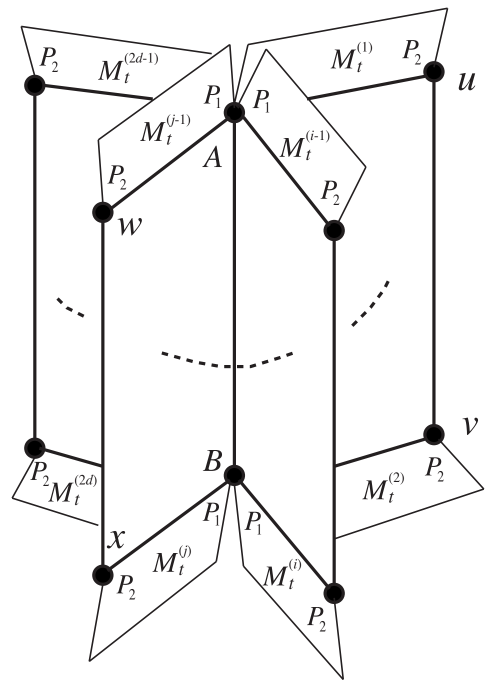

Figure 3: $M_d(t+1)$ is obtained from the juxtaposition of $2d$ copies of $M_d(t)$.

Thus, the sum of distances S_{t+1} satisfies the following recursion:

$$S_{t+1} = 2d\, S_t + \Delta_t. \tag{14}$$

where Δ_t is the sum over all shortest path length whose endpoints are not in the same $M_t^{(\eta)}$ branch.

To compute Δ_t, we classify the vertices of $M_d(t+1)$ into two categories: the two vertices with the largest degree (i.e., A and B in Fig. 3) are called *hubs*, while all other vertex are called non-hub vertex. Thus Δ_t can be obtained by adding the following path lengths that are not included in the distance between vertex pairs of $M_t^{(\eta)}$: length of the shortest paths between non-hub vertices, length of the shortest paths between a hub and non-hub vertices, and length of the shortest paths between hubs (for example, d_{uv}, d_{uB}, and d_{ux}).

Let us denote $\Delta_t^{\alpha,\beta}$ as the sum of all shortest paths between non-hub vertices, whose endpoints are in $M_t^{(\alpha)}$ and $M_t^{(\beta)}$, respectively. Thus, $\Delta_t^{\alpha,\beta}$ rules out the paths with endpoints at the connecting vertices belonging to

$M_t^{(\alpha)}$ or $M_t^{(\beta)}$. For example, each path contributing to $\Delta_t^{1,2}$ does not end at vertex u, v, A or B, and each path contributing to $\Delta_t^{1,4}$ does not end at vertex u, A, B or x. According to its value, $\Delta_t^{\alpha,\beta}$ can be split into three classes, where the three representatives are $\Delta_t^{1,2}$, $\Delta_t^{1,3}$, and $\Delta_t^{1,4}$, and the cardinality of the three classes are d, $d(d-1)$, and $d(d-1)$, respectively. Analogously, the length of the shortest paths between a hub and all non-hub vertices can be classified into two classes, while the shortest paths between hubs can be partitioned into three classes with path lengths equal to 1, 2, or 3.

Let Ω_t^{α} be the set of non-hub vertices in $M_t^{(\alpha)}$, then the total sum Δ_t is given by

$$\begin{aligned}\Delta_t &= d\Delta_t^{1,2} + d(d-1)\left(\Delta_t^{1,3} + \Delta_t^{1,4}\right) + 2d(d+1) \\ &\quad \sum_{j\in\Omega_t^2} d_{Aj} + 2d(d-1)\sum_{j\in\Omega_t^4} d_{uj} + \\ &+ (d+1)d_{uv} + d(d+1)\,d_{uw} + d(d-1)\,d_{ux}, \qquad (15)\end{aligned}$$

where $d_{uv}=1$, $d_{uw}=2$, and $d_{ux}=3$ are easily seen.

Having Δ_t in terms of the quantities of $\Delta_t^{1,2}$, $\Delta_t^{1,3}$, $\Delta_t^{1,4}$, $\sum_{j\in\Omega_t^2} d_{Aj}\sum_{j\in\Omega_t^2} d_{uj}$, and $\sum_{j\in\Omega_t^2} d_{uj}$, the next step is to explicitly determine these quantities. To this end, we classify non-hub vertices in $M_d(t+1)$ into two different parts according to their shortest path lengths to either of the two hubs (i.e. A and B). Notice that the vertices A and B themselves are not partitioned into either of the two parts represented as P_1 and P_2, respectively. The classification of vertices is shown in Fig. 3). For any non-hub vertex φ, we denote the shortest path length from φ to A, B as a, and b, respectively. By construction, a and b can differ at most by 1 since vertices A and B are adjacent. Then the classification function $class(\varphi)$ of vertex φ is defined to be

$$class(\varphi) = \begin{cases} P_1 & \text{for } a < b, \\ P_2 & \text{for } a > b. \end{cases} \qquad (16)$$

It should be mentioned that the definition of the vertex classification is recursive. For instance, class P_1 and P_2 in $M_t^{(1)}$ belong to class P_1 in $M_d(t+1)$, class P_1 and P_2 in $M_t^{(2)}$ belong to class P_2 in $M_d(t+1)$, and so on. Since the two hubs A and B are symmetrical, in the graph we have the following equivalent relations from the viewpoint of class cardinality: classes P_1 and P_2 are equivalent one to another. We denote the number of vertices in network $M_d(t)$ that belong to class P_1 as N_{t,P_1}, and the number of vertices in class P_2

as N_{t,P_2}. By symmetry, we have $N_{t,P_1} = N_{t,P_2}$, which will be abbreviated as N_t hereafter. It is easy to see that

$$N_t = \frac{|V(t)|}{2} - 1 = \frac{d(2d)^t - d}{2d - 1}. \tag{17}$$

For a vertex φ in $M_d(t+1)$, we are also interested in the smallest value of the shortest path length from φ to either of the two hubs A and B. We denote the shortest distance as this value by f_φ, and it can be defined as

$$f_\varphi = min(a, b). \tag{18}$$

Let δ_{t,P_1} (δ_{t,P_2}) denote the sum of f_φ for all vertices belonging to class P_1 (P_2) in $M_d(t)$. Again by symmetry, we have $\delta_{t,P_1} = \delta_{t,P_2}$ that will be written as δ_t for short. Taking into account the recursive method of constructing $M_d(t)$, we notice that the vertex classification follows also a recursion. Therefore we can write the following recursive formula for δ_{t+1}:

$$\delta_{t+1} = 2d\,\delta_t + d\,N_t + d. \tag{19}$$

Substituting equation (17) into equation (19), and considering the initial condition $\delta_0 = 0$, equation (19) is solved inductively

$$\delta_t = \frac{2d - 2d^2 - (2d)^{1+t} + d(2d)^{1+t} - dt(2d)^t + dt(2d)^{1+t}}{2(2d-1)^2}. \tag{20}$$

We now return to compute equation (15). For convenience, we use $\Gamma_t^{\eta,i}$ to denote the set of non-hub vertices belonging to class P_i in $M_t^{(\eta)}$. Then $\Delta_t^{1,2}$ can be written as

$$\begin{aligned}
\Delta_t^{1,2} &= \sum_{\substack{r\in\Gamma_t^{1,1}\bigcup\Gamma_t^{1,2}\\ s\in\Gamma_t^{2,1}\bigcup\Gamma_t^{2,2}}} d_{rs} \\
&= \sum_{\substack{r\in\Gamma_t^{1,1}\\ s\in\Gamma_t^{2,1}}} (d_{rA} + d_{AB} + d_{Bs}) + \sum_{\substack{r\in\Gamma_t^{1,1}\\ s\in\Gamma_t^{2,2}}} (d_{rA} + d_{Av} + d_{vs}) \\
&\quad + \sum_{\substack{r\in\Gamma_t^{1,2}\\ s\in\Gamma_t^{2,1}}} (d_{ru} + d_{uB} + d_{Bs}) + \sum_{\substack{r\in\Gamma_t^{1,2}\\ s\in\Gamma_t^{2,2}}} (d_{ru} + d_{uv} + d_{vs}) \\
&= 8N_t\delta_t + 6(N_t)^2.
\end{aligned} \tag{21}$$

Analogously, we find

$$\Delta_t^{1,3} = 8N_t\delta_t + 4(N_t)^2 \tag{22}$$

and

$$\Delta_t^{1,4} = 8N_t\delta_t + 8(N_t)^2. \tag{23}$$

Next we will determine other quantities in equation (15), with $\sum_{j\in\Omega_t^2} d_{Aj}$ given by

$$\begin{aligned}\sum_{j\in\Omega_t^2} d_{Aj} &= \sum_{j\in\Gamma_t^{2,1}} (d_{AB} + d_{Bj}) + \sum_{j\in\Gamma_t^{2,2}} (d_{Av} + d_{vj}) \\ &= 2\,\delta_t + 3\,N_t.\end{aligned} \tag{24}$$

Analogously, we can obtain

$$\sum_{j\in\Omega_t^4} d_{uj} = 2\,\delta_t + 5\,N_t. \tag{25}$$

Substituting equations (21), (22), (23), (24) and (25) into equation (15), we have the final expression for cross distances Δ_t,

$$\begin{aligned}\Delta_t &= 1 + 5d^2 + 4d(4d-1)N_t + 6d(2d-1)(N_t)^2 + \\ &+ 8d\delta_t[d + (2d-1)N_t] = \\ &= \frac{1}{(1-2k)^2}(1 - 4d + 5d^2 - 2d^3 + (1-d)(2d)^{2+t} + \\ &+ (5+2t)4^{1+t}d^{4+2t} - (7+2t)d^2(2d)^{1+2t}).\end{aligned} \tag{26}$$

Inserting equation (26) into equation (14) and using the initial condition $S_0 = 1$, equation (14) is solved inductively,

$$\begin{aligned}S_t &= \frac{1}{(-1+2d)^3}(-1 + 4d - 5d^2 + 2d^3 + 2^{1+t}d^{1+t} - \\ &- 7\cdot 2^{2t}d^{2+2t} + 3\cdot 2^{1+2t}d^{3+2t} - 2^{1+t}d^{1+t}t + \\ &+ 3\cdot 2^{1+t}d^{2+t}t - 2^{2+t}d^{3+t}t - 2^{1+2t}d^{2+2t}t + \\ &+ 2^{2+2t}d^{3+2t}t).\end{aligned} \tag{27}$$

Substituting equation (27) into equation (13) yields the exact analytic expression for the average distance of $M_d(t)$ as

$$\bar{D}(t) = (-1 + 4d - 5d^2 + 2d^3 + 2^{1+t}d^{1+t} - 7\cdot 2^{2t}d^{2+2t} +$$

$$
\begin{aligned}
+ &\quad 3 \cdot 2^{1+2t} d^{3+2t} - 2^{1+t} d^{1+t} t + 3 \cdot 2^{1+t} d^{2+t} t - \\
- &\quad 2^{2+t} d^{3+t} t - 2^{1+2t} d^{2+2t} t + 2^{2+2t} d^{3+2t} t) \\
/ &\quad ((-1+2d)(-1+d+2^t d^{1+t})(-1+2^{1+t} d^{1+t})).
\end{aligned}
\tag{28}
$$

References

[1] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.

[2] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.

[3] C. Song C, S. Havlin, H.A. Makse, Self-similarity of complex networks, Nature 433 (2005) 392–395.

[4] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.-L. Barabási, Hierarchical organization of modularity in metabolic networks Science 297 (2002) 1551–1555.

[5] M. Sales-Pardo, R. Guimerà , A.A. Moreira, L. A. Nunes Amaral, Extracting the hierarchical organization of complex systems, Proc. Natl. Acad. Sci. USA 104 (2007) 1522415229.

[6] A. Clauset, C. Moore, M.E.J. Newman, Hierarchical structure and the prediction of missing links in networks, Nature 453 (2008) 98–101.

[7] A.-L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks, Physica A 272 (1999) 173–187.

[8] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Pseudofractal scale-free web, Phys. Rev. E 65 (2002) 066122.

[9] F. Comellas, G. Fertin, A. Raspaud, Recursive graphs with small-world scale-free properties, Phys. Rev. E 69 (2004) 037104.

[10] J.S. Andrade, H.J. Herrmann, R.F.S. Andrade, L.R. da Silva, Apollonian Networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs, Phys. Rev. Lett. 94 (2005) 018702.

[11] J.P.K. Doye, C.P. Massen, Self-similar disk packings as model spatial scale-free networks , Phys. Rev. E 71 (2005) 01628.

[12] Z.Z. Zhang, F. Comellas, G. Fertin, L.L. Rong, High dimensional Apollonian networks, J Phys A: Math Gen 39 (2006)1811–1818.

[13] F. Chung, L. Lu, T.G. Dewey, D.J. Galas, Duplication models for biological networks, *J. Comput. Biol.* 10 (2003) 677–687.

[14] R.V. Solé, R. Pastor-Satorras, E. Smith, T.B. Kepler, A model of large-scale proteome evolution. Adv. Complex Syst. 5 (2002) 43-54.

[15] Z.Z. Zhang, S. Zhou, L. Fang, J. Guan, Y. Zhang, Maximal planar scale-free Sierpinski networks with small-world effect and power law strength-degree correlation, Europhys. Lett. 79 (2007) 38007.

[16] F. Comellas, Z. Zhang, L. Chen. Self-similar non-clustered planar graphs as models for complex networks. J. Phys. A: Math. Theor. 42 (2009) 045103.

[17] F. Comellas, A. Miralles. Modeling complex networks with self-similar outerplanar unclustered graphs. Physica A 388 (2009) 2227–2233.

[18] R. Ferrer i Cancho, C. Janssen, R.V. Solé, Topology of technology graphs: Small world patterns in electronic circuits, Phys. Rev. E 64 (2001) 046119.

[19] M.E.J. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.

[20] R. Diestel, Graph Theory, 3rd. edition, Springer-Verlag, Heidelberg, Germany, 2005.

[21] M.E.J. Newman, Assortative mixing in networks, Phys. Rev. Lett. 89 (2002) 208701.

[22] S. Valverde, R. Ferrer i Cancho,R. V. Solé, Scale-free networks from optimal design, Europhys. Lett. 60 (2002) 512–517.

[23] R.V. Solé, S. Valverde, Information theory of complex networks: On evolution and architectural constraints, Lect. Notes Phys. **650** 189-207 (2004).

[24] H. Jeong, S. Mason, A.-L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42.